

Embedding Open Science in Reality

John Wood[†]

ATTRACT Project Office, Development & Innovation (IPT-DI), CERN, CH 1211 Geneva 23, Switzerland

Keywords: Open Science; Closed science; Data

Citation: Wood, J.: Embedding open science in reality. *Data Intelligence* 3(1), 106-115 (2021). doi: 10.1162/dint_a_00083

chinaXiv:202211.00406v1

ABSTRACT

The Open Science (OS) movement has achieved extraordinary results in very few years. In this paper I argue it is now necessary to embed OS in the wider ecosystem of research and innovation, acknowledging some of the outstanding issues that need to be resolved as it beds down into the way research is done in the future. By sticking to a purest approach to OS its impact and current momentum may be lost. Digital technologies and global connectivity have ensured that OS is here to stay and will continue to expand its influence in the future. However, OS cannot stand aloof from what is the reality of what is happening elsewhere otherwise it will do a disservice to itself and the challenges facing the world.

1. A HISTORICAL PERSPECTIVE

The idea of Open Science (OS) (specifically Open Data (OD)) is not new although it is only in the past 15–20 years that it has become a buzz word for new ways of doing science fostered by the massive increase in the capacity to exchange knowledge in a blink of an eye electronically across the world. The collection and analysis of information has been around since written records began. These were collected to enforce taxes, plan food supplies and mobilize armies among other things. How open this information was dependent on who you were since the information was lodged largely in one central location and access was limited to those with a need to know or who could actually read. In the UK in 1086, just under 1000 years ago, a complete survey of England was undertaken at the request of William the Conqueror and each village's resources were tabulated in a surprisingly short time. The books (three in total) where these details were recorded are known as the "Doomsday Book." Originally written in Latin, the details are available to all today in an English translation [1]. There are four mentions of the village where I live which now has about 46 houses widely dispersed across the countryside. The first entry tells me that one Tovi who was a

[†] Corresponding author: John Wood (Email: johnvwood@gmail.com; ORCID: 0000-0003-3974-5619).

priest looks after “half a hide and enough woodland to house 30 pigs” on behalf of the Bishop of Bayeux. A “hide” was a unit of land sufficient to sustain a family at the time. Another resident, Guthmund, has woodland for 20 pigs but this woodland is owned by the Bishop of Coutances. While this information is of intense interest to social historians and those that live in the village it is unlikely to excite anyone else. It is OD, yet it says nothing about the present state of the village. One might assume that Tovi as a priest lived near the church (which is about 1000 years old); however today there are no houses near the church since Black Death wiped out most of the village in 1348 so we have no idea where either Tovi or Guthmund lived or whether their woods survive today. The reason for bringing this up is to show that while the information is open and accessible it is both time dependent and there is a need for metadata to interpret terms which are not used in modern speech. While the information was true at the time of recording it is no longer true of the current situation.

chinaXiv:202211.00406v1

Bringing ourselves to the present, the term “open science” has evolved enormously since the embryonic stages of open access and institutional repositories some 20 years or more ago. Actually, the reality is that the idea of “open science” and the tensions caused in its implementation have been around since science in the modern era began. Isaac Newton did not want to share his ideas and only published to show he had thought of things first before those who were more open. In fact, not only did he not want to publish, he demanded information from others to support his calculations [2]. This is an early example of “what is yours is mine and what is mine is my own.” While Isaac Newton had his problems with authorities and especially the church [3] these did not intrude on his science where he used information from others freely without hindrance although he wrote scathing letters to those who would not give him the information he wanted. However, others like Galileo Galilei published their data openly and fell foul of those in the church who had vested interests and wanted to prevent him sharing his evidence (e.g., [4]). Even today there is a tension between scientists who are intensely secretive about their work and those that collaborate willingly. In many ways, personal accolades and promotion procedures in universities foster secrecy and it is only in subjects such as particle physics and astronomy where openness has been the predominant culture for some time although other areas are catching up. While this is the situation at the personal level, it is the same at both the corporate and national level. Having spent a lifetime as an academic I have found that as a general rule, one third of academics willingly work together and want to share information, one third will do it if there is grant money available and one third feel it is an intrusion on their need to be academically free and work alone. While funders may or may not insist that all publicly funded research is publicly available, the ability to police such policies is very difficult for individuals although easier where groups from different disciplines and backgrounds need to share information. The lesson from this is that while it is easy to make pronouncements the reality is much less clear and it is likely that we will have to live with a mixed approach in the future. Sometimes holding to a rigid doctrine can defeat the very purpose of that doctrine. The question is how to encourage OS without killing originality.

2. A TAXI EXAMPLE

Here is a further example to demonstrate how closed and OS cohabit. Although once common throughout Europe, today the greatest concentration of trade guilds (livery companies are the official name in the UK and most carry the title “Worshipful Company” before the trade they represent) are in the City of London. (Just for the record, the “City of London” is just the square mile at the heart of London which has an independent governance structure from the rest of England^①). Many of them were founded almost 1,000 years ago but some are still being created. Of the 110 that exist today number 104, which was granted livery status in 2004 is the “Worshipful Company of Hackney Carriage Drivers” or, in other words, those that drive the black cabs around greater London (not just the City).

However, behind all these companies is what looks like closed information called their “mystery” which is a corruption of the French word “metier” or trade. Previously they kept this information to themselves and only by a complicated process would admit new members to share in this information. In the case of the Hackney Carriage Drivers this information is called “the knowledge” which all cab drivers have to show they have learnt before they receive a licence and are allowed to drive a taxi in London. While at first, this may seem a case of “closed information,” in reality this information is freely available but takes a long time and patience to learn involving walking or cycling around the best ways of getting around London. Incidentally it is found by MRI (Magnetic Resonance Imaging) scanning that those who have absorbed this knowledge have a larger area of the brain (memory) than the average person. This is followed by an oral test administered by an independent body on the best way to get from random places. They are not allowed any electronic device for this. Once they obtain their licence they can, if they wish, apply to join the livery company and join the “mystery.” After all this effort it is little wonder the cab drivers are not happy with companies like Uber whose drivers just use an electronic device to find their way around and do not have the “knowledge”. The satnav gives “open information” but does not necessarily know the little nuances that those with the “knowledge” have. Yet, in reality, both live side by side and the mixture of closed and open information on the same subject manages to coexist albeit it with many tensions.

3. TENSION OR HARMONY BETWEEN OPEN AND CLOSED SCIENCE

The reason for bringing up this example is that it starts to open up the complex nature and tension between “open science” and “closed science.” What appears to be a clear distinction is not so clear in practice especially when commercial advantage and livelihoods are at stake. What appears open can in fact be closed and vice versa. So, is it idealistic to think about a truly OS environment or will it, in reality be a mixed economy? Or to use the words from the FAIR approach to data but to widen it further to encompass the whole of OS: “open as possible and as closed as necessary.” This is an issue which currently faces the ATTRACT project which involves Eiroforum (The European Intergovernmental Research

^① See https://www.youtube.com/watch?v=LrObZ_HZZUc for an amusing explanation.

Organizations Forum) members, universities and trade bodies^② whereby very early stage ideas in sensor and instrumentation technologies are funded by public (EC) money but are expected to transfer to private funding as the projects progress to near market. The strap line for the ATTRACT project is “Open Science for Open Innovation” with the sub-text of getting Europe back to work. What has been surprising in the first phase of this project is the number of private investors wanting to become involved much earlier than initially expected. A few reasons are given for this. One is that they feel that inventors probably do not realize there are easier markets to penetrate than at first were thought of. Secondly, they are looking for people who can drive a product through from inception and finally they like the brand image of ATTRACT with the institutes behind the project. However, in mixing public and private money the concept of OS is put under strain. The situation is not resolved but will be the subject of intense discussions between ATTRACT, the European Investment Council, regional and private funders in the coming years.

4. A VIEW FROM THE OS PLAINS

Even if we come back to the clear hinterland for OS without some of the above fuzzy boundaries, things are far from clear. OS is widely regarded as a catch all term covering everything from OD, open access through to Citizen Science (CS) and ultimately leading on to tangible outcomes via open innovation. The publication *Progress of Open Science: Towards a Shared Information Knowledge System* [5] produced by the Open Science Policy Platform of the EC states:

“Even though the tools and technology to enable Open Science has been available for almost two decades, progress has been slower than anticipated and there remain real obstacles to overcome. Notably, there is a disparity in progress and motivation among different disciplines and institutions, among different actors and organizations, and among researchers at different stages of their career. This is compounded by a lack of policy alignment across local, regional, national and international jurisdictions, such as across Member States, and no clear legal or regulatory framework, often associated with insufficient cost/benefit analysis of Open Science requirements.

Open Science for its own sake has never been the goal. While a focus on Open Science as a mechanism must be emphasized in any transition, Open Science must ultimately be embedded as part of a larger more systemic effort to foster all practices and processes that enable the creation, contribution, discovery and reuse of research knowledge more reliably, effectively and equitably. Research cannot be ‘excellent’ without such attributes at its core.”

Those of us who had a clear vision of what “open science” was all about from the start tend to come from a research background which is mainly funded by public funds. It is here that much of the policies

^② <https://attract-eu.com/>

and actualities of OS predominate including the development and activities of Plan S, FAIR principles, OpenAire, European Open Science Cloud (EOSC) and work of CODATA and the Research Data Alliance. A key idea that attracts political support is that OS allows the reuse of data and thus increases the return on investment. However, there are obvious societal benefits such as the current work of a number of groups sharing data on the COVID-19 virus. The Research Data Alliance has eight working groups sharing information on different aspects of COVID-19 from social sciences through to modelling and legal and ethical aspects. A good example of how this works in practice is found in two of the projects supported by ATTRACT. A small start-up company (AquAffirm) has developed a range of sensors for detecting ultra-low levels of contaminants in drinking water from bored wells (initially arsenic and fluoride) that can lead to dramatic consequences in the development of cancer over the long term [6,7]. The underlying software package to monitor and consolidate the information from these sensors has now been used in COVID-19 research. Details can be found at www.covidsim.org which shows the epidemic trajectory and healthcare demand simulations based on epidemiological modelling algorithms developed in conjunction with Imperial College London. This tool enables government healthcare officials, journalists and researchers from low- and middle-income countries to deploy advanced epidemiological prediction tools. The Web-based tool informs economic and political decisions concerning intervention/restriction strategies and related resource allocation. A further development of this software now enables civil engineers to optimize the routing of services such as water supplies, effluent in large sustainable civic structures. This is a clear case where OS leads to wider benefits.

Other notable tangible results other than purely technical have been in the sharing of data for materials developments, rice and wheat for sustainable food and many other areas. Much focus has rightly looked at the reliability, traceability, availability, etc. of data and organizations such as GO FAIR are helping to ensure that such data are truly open and useable.

5. DROWNING IN DATA

As the opening paragraph of this paper shows, data can fulfil all the principles but still do not give a full picture of the truth. Much is made of scientific integrity among researchers. Yet the number of retracted papers and examples of scientific misconduct continues to grow. The first European Research Area report entitled “Preparing Europe for a New Renaissance” [8] argued for a “social contract” along the lines of the oath taken by new medical doctors, which integrated scientific excellence paired with social awareness and responsibility including ethical, social and economic dimensions. Taking science and especially OS out of context as if it is divorced from the bigger picture has to be resisted. It is two sided and politicians and policy makers have to treat OS in a way that acknowledges its contribution in a truthful way. In the pandemic crisis the politicians hid behind the statement that they will be guided by the science. Other statements such as “science says” are also banded about. Many times the science does not see the whole picture and in pushing the claims of OS beyond what it can deliver it provides the opportunity to discredit its considerable contributions.

6. NEW WAYS OF WORKING AND ACCOUNTABILITY

In the European Commission's report "Riding the Wave" [9] which was the original report that kick started the Research Data Alliance from the European perspective, there are a number of potential scenarios emphasizing the need for trust between researchers as they share information. One of these involved how things might look for a research student in the future.

"Roger is working on an international Ph.D. It's a relatively new program, in which the student applies to become a member of an international team working on a big problem that affects all people. His group is comparing many forms of non-verbal communications between cultures. It has several hundred members and his university tutor is one of the nodal points contributing expertise in "synergistic communication between biological components". Others in the network are using archaeological evidence to study communications between ancient Mesopotamian and Hellenic cultures; some are studying computer-computer interactions between different systems; yet more are studying communications in refugee camps. Each node contributes to the whole. Results are communicated as they happen, and there are daily virtual-presence planning sessions. Roger has to sign a contract not to misuse data or contribute anything that is not for the common good—such as externally sources information that he has not thoroughly checked for provenance."

While this is a purely hypothetical case it does highlight two important issues. The first is for a contract of behavior and the second is the nature of research in the OS future. At the moment we have a bottom up anarchy and while we should avoid top down regulation, this anarchy does need to self-organize and there are groups in the Research Data Alliance (RDA) and elsewhere who are tackling these issues. It is to be hoped that a common ground will not only be agreed but be taken up worldwide as common standards. In this it has a number of similarities with the telecoms industry. Yet most universities still operate and sustain systems that still favor an individualistic approach to research. While each university and research organization is independent and should remain so there is a need for a system of (I do not want to use the word metrics which sounds very prescriptive) criteria that could be used and weighted according to the needs of the organization. Unfortunately, the emphasis on university league tables acts contrary to this position. In the OSPP report [5] it recommends that this is an issue that any future OSPP with the RDA might take up. In doing this it is necessary to be mindful of the way many, mainly science and technical, universities are changing their teaching approaches to involve holistic approaches to student-student learning which rely heavily on many of the principles of OS. While there are many examples in the USA of institutions taking this approach such as Olin College in New York, it is encouraging to see this approach being taken up in Europe and Asia. A good example is how Nanyang Technical University in Singapore has embraced this approach in its undergraduate teaching program.

7. EDUCATION

Embedding OS both as a methodology and as culture into all levels of education is essential, starting from primary school right up to research training. At one end of the spectrum is the necessity for all citizens to appreciate how to interrogate data away from the hype of politicians and the headlines of the media. At the other end is the need for professional data scientists who are part of the research process with a clear career path. The Edison project supported by the European Commission^⑤ was an attempt to create a total training package which was possibly too restrictive. Universities are now offering Data Science courses often linked with computing departments or business schools. CODATA (Committee on Data for Science and Technology) with TWAS (The Third World Academy of Sciences) and the RDA are hosting training courses for academics in developing countries. A number of employment agencies are now seeing data scientists as a specific profession. As an example, Prospects, a small specialized recruitment agency promotes the role of the data scientist on their website^⑥. They list eight business areas, including academia, as examples of where there are opportunities. Here is their introduction to the topic which clearly shows that they believe there is a seamless link between OS and open innovation.

Data scientists turn raw data into meaningful information that organizations can use to improve their businesses

Organizations are increasingly using and collecting larger amounts of data during their everyday operations. From predicting what people will buy to tackling plastic pollution, your job is to use data to find patterns and help solve the problems faced by businesses in innovative and imaginative ways.

You'll extract, analyze and interpret large amounts of data from a range of sources, using algorithmic, data mining, artificial intelligence, machine learning and statistical tools, in order to make it accessible to businesses. You will then present your results using clear and engaging language.

Data scientists are in high demand across a number of sectors, as businesses require people with the right combination of technical, analytical and communication skills.

The other main area which is growing rapidly is that of CS which is now reaching into all areas of research. The great danger here is the need to ensure quality. This is articulated in the LERU report [10]. In this paper they say:

"We distinguish three important trends:

^⑤ <https://edison-project.eu/edison/edison-project/>

^⑥ <https://www.prospects.ac.uk/job-profiles/data-scientist>

- 1). *Increasing coordination and collaboration between CS practitioners from different fields, which leads to sharing procedures and best practices, and to the creation of networks and associations.*
- 2). *Emergence of platforms that support a variety of CS projects, creating broader public awareness and encouraging a greater retention of volunteers.*
- 3). *Expanding the role played by citizens in the projects beyond simple tasks to include greater participation in all phases of the research process from conceptualization to publication."*

The report goes on to make detailed recommendations for researchers, universities, funders and policy makers. There are now several university courses available either as part of undergraduate programs or as standalone that try to teach the principles behind OS. Funders are now looking at how they can fund high quality CS that passes the normal peer review process.

chinaXiv:202211.00406v1

8. MAKING IT PAY

Policy makers and funders have largely bought into the fact that OS is good value for money yet the various analyses that have been undertaken are not that rigorous and maybe it is not worth pursuing much further given OS is largely accepted. A recent paper by Fell [11] looks at methodologies for assessing the economic impact which, unsurprisingly argues for further work on agreed metrics. Probably more effective is to see how OS impacts on open innovation [12]. Unfortunately, this was not part of the remit of the OSPP report and it may be that initiatives such as the ATTRACT project are needed to give solid evidence that there is a linkage. Although researchers may say that doing OS for its own sake is sufficient justification, sooner or later the funders will be asking the question and it is expedient that the community undertake studies in this area before being asked. The rise of neo-nationalism in many countries coupled with the lack of freedom to be open in others means that there will be questions regarding the underlying principles of OS by certain politicians that could cause a negative backlash.

9. FINAL COMMENTS

OS has been fantastically successful so far, aided both by the developments in computing power and in globalization. It is now time for a reality check to make sure it is firmly embedded in the wider research/scholarship/innovation ecosystem. There is a long way to go and some compromises will be necessary. It has taken over 20 years for open access to be largely accepted and initiatives like Plan S are now official policy with many funders. OS has been largely bottom-up led which is its strength. When I first presented the idea of OS to the EU's Competitiveness Council I argued that it should not be regulated and be left free to run its own course. Unfortunately, some boundaries have to now be set and the OS community needs to ensure they are not restrictive. In many areas, reality has to be faced and compromises will be necessary. In some ways, the exciting phase of development is over and now begins the drudge of taking things forward

positively and embedding OS as the norm. On a personal level I feel very privileged to have been in at the start of the process. History will judge our achievements.

ACKNOWLEDGEMENTS

I am grateful to Kostas Glinos for his comments and also for inviting me to chair the first high-level group to look at the future of scientific data in Europe that led to the creation of the Research Data Alliance (RDA).

REFERENCES

- [1] Williams, A., Martin, G. (eds.): *Domesday book*. Penguin Books, London (1992)
- [2] Westfall, R.: *Never at rest—a biography of Isaac Newton*. Cambridge University Press, Cambridge (1980)
- [3] Iliffe, R.: *Priest of nature*. Oxford University Press, Oxford (2017)
- [4] Livio, M.: *Galileo and the science Deniers*. Simon and Schuster, New York (2020)
- [5] EU Open Science Policy Platform (#EUOSPP) Final Report. Available at: <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-policy-platform>. Accessed 4 January 2021
- [6] Wood, M.: *Arsenic—Detecting invisible poisons*. Chemistry Review 30(2), 7–9 (2020)
- [7] Muller, M., et al.: Using random testing in a feedback-control loop to manage a safe exit from the COVID-19 lockdown. Available at: <https://arxiv.org/pdf/2004.04614.pdf>. Accessed 4 January 2021
- [8] Burgelman, J.C.: Preparing Europe for a new renaissance: The vision and the proposed actions of ERAB. Available at: https://www.researchgate.net/profile/Jean_Burgelman/publication/265118746_Preparing_Europe_for_a_New_Renaissance_the_vision_and_the_proposed_actions_of_ERAB/links/54ac10260cf25c4c472fe605/Preparing-Europe-for-a-New-Renaissance-the-vision-and-the-proposed-actions-of-ERAB.pdf. Accessed 4 January 2021
- [9] Riding the wave. Available at: https://ec.europa.eu/eurostat/cros/content/riding-wave_en. Accessed 4 January 2021
- [10] Wyler, D.: Citizen science at universities: Trends, guidelines and recommendations. Available at: <https://www.ieu.org/publications/citizen-science-at-universities-trends-guidelines-and-recommendations>. Accessed 4 January 2021
- [11] Fell, M.J.: The economic impacts of Open Science: A rapid evidence assessment. Publications 7(3), 46 (2019)
- [12] Chesbrough, H.: *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Review Press, Cambridge (2006)

AUTHOR BIOGRAPHY

Professor John Wood CBE, FREng, initially a metallurgist, is currently chair of the ATTRACT advisory board based at CERN in conjunction with other EIROforum members seeking to develop future disruptive technologies. He is currently chair of the Ciemat Advisory Board in Madrid (the Spanish Energy Research Institute). He has held academic posts at the Open University, Nottingham University and Imperial College (as Principal of the Faculty of Engineering). From 2001–2007 he was chief executive of the Council for the Central Laboratory of the Research Councils. Currently he holds visiting chairs at Imperial College, Brunel, Cranfield and Liverpool Hope Universities and has honorary doctorates from universities in Romania, Ghana and India. He is a fellow of the Royal Academy of Engineering and was made a CBE in 2007 for “services to science.” In 2010 he was made an “Officer of the Order of Merit of the Federal Republic of Germany.” He is an honorary citizen of the city of Cluj-Napoca in Romania. He remains an adviser to several worldwide universities, governments and the European Commission. He is a trustee of the Episcopal University of South Sudan and a council member of Brunel University. He is a non-executive director of a number of companies including being chair of AquAffirm a company specializing in detecting arsenic and other dangerous substances in drinking water. After his academic career he became Secretary-General of the Association of Commonwealth Universities until 2017 and is now a Commissioner for the Commonwealth Scholarships Commission.

ORCID: 0000-0003-3974-5619